DOCUMENT RESUME

ED 279 181 FL 016 489

AUTHOR Thavenius, Cecilia
TITLE Pronominal Chains in English Conversation.
PUB DATE 84
NOTE 12p.; In: Ringbom, Hakan, Ed,; Rissanen, Matti, Ed. Proceedings from the Nordic Conference for English Studies (2nd, Hanasaari/Hanaholmen, Finland, May 19-21, 1983); see FL 016 498.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Coherence; Cohesion (Written Composition); Comparative Analysis; Discourse Analysis; *English; *Language Patterns; Language Research; Language Usage; *Oral Language; *Pronouns; Semantics; Standard Spoken Usage; Syntax; *Word Frequency; *Written Language
IDENTIFIERS Referents (Linguistics); *Spontaneous Speech

ABSTRACT

A study of the functions and structure of referential third- person pronominal chains in spontaneous English conversation as compared to those in written English is discussed. The study found the frequencies of these pronouns and of their chains to be generally much higher in the spoken corpus than in the written corpus. In the spoken corpus, "it" chains were the most common, "she" chains were the least common, and most were endophoric. In the written corpus, "they" chains were the most common and "she" chains the least common, and all were endophoric. Spoken chains were longer than written chains and contained more tokens, which can be explained at least in part by factors in the context of situation. It was also found that chain length was closely related to the number of pronouns per chain, and that the range for chain length was much larger in conversation. The study concludes that differences in the use of pronominal chains in conversation and writing are due to differences in referential strategy: while a speaker in conversation makes pragmatic word choices and has no real wish to vary his words, a writer makes stylistic choices and aims at variation in the written product, which acts as a constraint on the use of pronominal chains. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
N. Enkvist
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED279181

Cecilia Thavenius
Department of English
University of Lund

PRONOMINAL CHAINS IN ENGLISH CONVERSATION

INTRODUCTION

Working with a corpus of spoken English such as the London-Lund corpus is rewarding in many ways. It makes one realize that spoken language differs from written language in more respects than one may have foreseen. It makes one perceive concepts and problems which one expected, and also which one did not expect. It may, for instance, also reveal that certain linguistic patterns have functions in conversation that are important not only for the cohesion and coherence of the text, but also for the interaction process. This became obvious to me when I made a study of the use and functions of third-person referential pronouns in face-to-face conversation (Thavenius 1983). Among other patterns, I examined the pronominal chains, that is sequences of two or more coreferential pronouns with no intervening noun.

When studying the chains in the London-Lund corpus, I discovered that they had functions in conversation which would not occur in written production. Halliday (1980:11) argues that text is "a continuous process of semantic choice". This statement is relevant to both written and spoken production. I would, however, like to add that written text may be seen as a continuous process of stylistic choice, whereas conversation (the kind of spoken language I deal with) may be seen as a continuous process of pragmatic choice. It is thus a basic assumption that this difference may be realized as different patterns in conversation and written text.

The aim of the present study is to discuss and examine the functions and the structure of third-person pronominal chains in spontaneous conversation, and then, to some extent, compare them with those in written production. The chains have been restricted to those containing instances of the pronominal groups IT, THEY, HE, and SHE. IT represents in this study not only 'it' (ie the word forms it, its, itself) but also the demonstratives 'this' and 'that' in head function. In this function 'this' and 'that' refer in their own right just like 'it'. As determiners 'this' and 'that' have not been included, as

FL016489

BEST COPY AVAILABLE

they have no independent referring function. THEY denotes 'they' (they, them, their, theirs, and themselves) as well as 'these' and 'those' in head function. HE denotes the word forms he, him, his, himself; SHE denotes she, her, hers, and herself.

It should also be noted that these pronouns are only relevant for this study when their primary function is referential. This means that pronouns in syntactic constructions like extraposition and cleft sentences, tag questions, and expressions concerning time, weather, and distance have been left out.

Pronominal chains have functions which are relevant both to written and spoken production. A chain is produced if there is a reason for referring more than once to something; if pronouns convey the whole meaning of the referent; if there is a wish to avoid repetition; and if there is no wish to vary the words with coreferent nouns. The pronouns then have a substitute function and contribute to a meaningful simplicity, which may, however, differ in extent between spoken and written language.

The chains also function on a semantic level where they build threads of meaning and form part of the semantic and reference system of the text. The pronouns in a chain contribute both to the cohesion of the text as overt linking signals, and to the coherence by achieving consistency with the situational context. The referent of a chain may be found in the spoken or written text itself; the reference relation is then endophoric. It may also be exophoric, that is there is no mentioned referent in the text and the pronouns refer directly outside the text, either to the physical situation or to the complex of cultural, general and specific knowledge of the world. In Thavenius 1982 and 1983, it was shown that exophoric pronominal reference was, on the whole, very infrequent in conversation (c 10%). In written language, exophoric reference should be still more infrequent, as direct reference to the physical situation can occur very seldom. It should perhaps also be mentioned that the fact that there is an endophoric relation between the pronouns in an exophoric chain has been disregarded in this study. All coreferent pronouns included in a chain refer to the same referent and that is why only the exophoric relation to the original referent has been taken into account.

There are, however, functions which are relevant in conversation but not in written production. A writer has a wish to vary his language. When he refers to the same referent several times, he tries to avoid repetition of the same word, that is he aims at stylistic variation. Not only may he avoid using the same noun several times, but also using too many pronouns.

This stylistic awareness obviously varies between different types of written texts as well as between different writers. I think, however, that there is, on the whole, an intentional attempt at variation, which is typical of written production, and which may function as a constraint on the use of pronominal chains in writing.

In conversation, on the other hand, it seems that pronominal chains play an important part in the communication process. Here, there is no constraint on using many pronouns as long as they convey the right meaning to the listener. The chains are, in fact, also essential to the process of interaction. As a speaker can never attain full knowledge of all the components in another speaker's discourse model, he uses a heuristic technique in order to check that the other speaker understands him. This means that he may go on adding pronouns to a chain either until the other speaker signals that he has not understood the meaning of the pronoun, or, until he himself feels that an elucidation is mandatory to avoid misunderstanding.

Pronominal chains contribute in fact to the cohesion and the coherence of conversation in a special way. When a chain is uttered by one speaker only, the result is a cohesive relation between the pronouns and the textual referent, if there is one. Thus the chain ties the parts of the utterance together and enhances the coherence of the conversation. Moreover, when a chain is produced by two or more speakers, the speakers' utterances or turns are tied together by means of the pronouns that form the chain. The speakers' utterances result in a coherent unity, to a large extent brought about by this kind of discourse cohesion. Discourse cohesion is a structural device of tying utterances together in which pronominal chains play an important part.

Chains may also function as a strengthening factor in the process of interaction. A speaker who initiates a chain gets a very strong response from the other speaker, when the latter adds a pronoun to the chain. In this way, the chain is built up by both speakers and contributes to the interaction. The following is an example of this interweaving process:

A I like that one in the corner● very +much●+
C +/m/●+ - yes● yes that's rather fine
A yes that's very nice●

(Svartvik & Quirk 1980:207)[1]

A introduces the referent, ie that one in the corner, which is referred to by C, and then C's reference is supported by A. It is obvious, however, that the

longer a chain is and the more pronouns it contains, the more we can expect several speakers to participate in it. Also, the shorter it is the less we can expect more than one speaker to participate in it.

My hypothesis was that there will be more chains in conversation than in writing, since chains seem to have more functions in spoken interaction than in written production.

The structure of chains can be discussed from two different but related aspects, namely chain length, i e the distance between the first and the last pronoun, and number of pronominal tokens per chain. The number of tokens can vary for many reasons, one of which is in fact length of the chain. A long chain will normally contain more tokens than a short chain. There is of course the occasional case when a very long chain, for some reason, contains only two tokens, but this is presumably exceptional. In Thavenius 1983, it was shown that there was a significant positive correlation between chain length and number of tokens per chain. Kroch & Hindle (1980:248), who made a study of pronominal chains, measured length in number of tokens only; they did not deal with distance at all. We may hypothesize that there will be longer chains in conversation than in writing, since there are so many verbal and non-verbal clues in conversation, contributing to the listener's understanding and to his possibilities of tracing referents, clues which are not present in a written text. Likewise, the chains probably contain more tokens in conversation.

A crucial factor for the frequencies and the structure of the chains is, however, the extra-textual context. There are so many contextual variables which may cooperate in a conversation and thus influence the frequencies and lengths of chains, for instance, that it would be extremely surprising if the chain patterns were not different in conversation and written text. Variables like conversation topic and relationships between the speakers may no doubt be decisive.

To sum up: there were reasons to expect that the chain frequencies are higher in conversation than in written English; that exophoric referents are more frequent for chains in conversation than in written text; that the chains contain more pronouns in conversation than in writing; and finally that the chains are longer in conversation than in writing.

MATERIAL

The chains have first been examined in a corpus of spoken English, and then the results have been compared with a small sample of written English. The

corpus of spoken texts I have used is the London-Lund Corpus of Spoken English (LLC). This corpus comprises 'texts' of spoken language ranging from face-to-face conversation, recorded surreptitiously (ie with a hidden microphone), to prepared but unscripted oration. The texts consist of c 5000 words each and have been subject to a prosodic and paralinguistic analysis at the Survey of English Usage, University College London. The material was taken over in 1975 by the Survey of Spoken English at Lund, where it was transferred to magnetic computer tape.[2]

For this study, eight texts were chosen from the category of spontaneous face-to-face conversations, surreptitiously recorded.[3] In all, these eight conversations amount to c 40,000 words. The speakers are all educated British people.

For a comparison with pronominal chains in written text I chose five texts, each of 2000 running words, from the Lancaster-Oslo-Bergen Corpus of British English (LOB).[4] The texts were all chosen from categories of informative prose, the kind of written language one might expect the speakers in the spoken corpus to read and produce. The written material is of limited size, only one fourth of the size of the conversation material, but it may still show whether the use of chains tends to differ between spoken and written language.

RESULTS

To begin with, the frequencies of third-person pronouns were on the whole (ie regardless of chain membership) much higher in the spoken than in the written corpus. Table 1 shows the frequencies in the two corpora per 10,000 words. The size of the spoken corpus is 40,000 words, but for easier comparison with the written corpus the former has also been expressed in number of pronouns per 10,000 words.

Of the pronouns as much as 69% were chain tokens in LLC, and 44% in LOB. This means not only that more pronouns were used in conversation than in writing, but also that the proportion of pronouns in chains was higher in conversation.

Table 1. Frequencies of pronouns per 10,000 words. In brackets the absolute frequencies in LLC, ie per 40,000 words.

	LLC	LOB
Total	591 (2365)	328
In chains	406 (1624) (69%)	143 (44%)

Table 2. Chain frequencies per 10,000 words.

	LLC	%	LOB	%
IT	52 (208)	49.7%	12	23.1%
THEY	26 (104)	24.8%	23	44.2%
HE	19 (78)	18.6%	16	30.8%
SHE	7 (29)	6.9%	1	1.9%
Total	104 (419)	100%	52	100%

Table 2 shows that the chain frequencies were higher in conversation. The total number of chains (per 10,000 words) in LLC was 104, compared with 52 in LOB. There were thus twice as many chains in conversation. In LLC, the IT chains were most common and SHE chains least common. In LOB, THEY chains were most common, and SHE chains least common. Why IT chains were most common in LLC can be explained by the range of functions this pronoun has; but why IT chains were so infrequent in LOB is difficult to explain. One reason might be the informative and strictly topic-determined character of the texts, but an explanation of the low frequency of IT chains in LOB re-

quires a more extensive study.

It was suggested that exophora would be even less frequent in writing than in conversation. To test this hypothesis I studied the chain referents in the two corpora and found (see Table 3) that 89.3% of the chain referents were endophoric and that 10.6% were exophoric in LLC. In LOB, all chain referents were endophoric.

Table 3. Proportions of endophoric (N) and exophoric (X) referents for the chains.

	LLC		LOB	
	N	X	N	X
IT	85.6%	14.4%	100%	-
THEY	84.6%	15.4%	100%	-
HE	94.9%	5.1%	100%	-
SHE	100%	-	100%	-
Total	89.3%	10.6%	100%	-

Even if it might be argued that the written corpus I used was very limited in size, 10,000 words with only 52 chains, there is a clear tendency to predominating endophoric reference in both channels that cannot be denied, and it is in fact stronger in the written texts. It is clear that the current topic in the text, depending on text type, can influence the use of exophoric chains, but there are not many possibilities in writing to refer directly outside the text.

To test the hypothesis that the chains contain more tokens in conversation than in writing, the number of pronouns per chain was counted. In the conversation corpus this was done automatically by computer, since that study was part of a more comprehensive one with all information stored on magnetic tape (cf Thavenius 1983). For the written sample the count was performed manually.

Table 4 shows the mean number of pronouns per chain, and also the range in LLC and LOB, respectively.

Table 4. Number of pronouns per chain. Brackets denote that there was only one SHE chain in LOB.

	LLC	LOB
IT	2.9	2.3
THEY	3.6	2.6
HE	6.4	3.3
SHE	5.1	(2)
Total mean	4.5	2.6
Total range	2-37	2-8

It was found that no chain contained more than eight pronoun tokens in the written corpus, whereas the chains could contain up to 37 pronouns in conversation. It is obviously easier to keep track of pronoun referents in conversation. There was, however, no IT chain in the conversations that contained more than eight pronouns. The word it has many referential functions, and can for instance refer not only to noun phrases but also to the contents of whole text passages, a function which the other third-person pronouns lack; it also has functions which are not primarily referential, and the result is that the speakers in a conversation are exposed to so many instances of the word it that it is more difficult to keep track of referents for IT than for the other pronouns. In both LLC and LOB, HE chains contained more tokens than the other pronominal chains.

Kroch & Hindle (1980) found that pronoun chains in speech may contain up to 32 pronouns, whereas no chain in their written sample contained more than six pronouns. Although their study was made on American English and the sample was even smaller than my written corpus, the tendency is the same; chains in speech contain more pronouns than chains in writing.

As it has already been shown (cf Thavenius 1983) that there is a positive correlation between number of pronouns per chain and chain length, a study of chain length cannot add much to what has been said above. We may, however, wonder how long the chains generally are. As a matter of fact, they are usually fairly short both in conversation and in writing. In the written corpus, chain length was measured in terms of number of words between the first and the last pronoun in the chain. In the conversation corpus the

chains were measured in number of tone units, that is the basic unit of analysis for LLC (cf Crystal 1976:207ff). One tone unit contains on the average 5.3 words (cf Quirk et al 1964). To be able to make a rough comparison of chain length in the two corpora, the number of tone units was multiplied by 5.3. The means for chain length are uninteresting, as some of the individual chains were very long and caused a certain skewness. Instead, the range and the proportions of chains that were between two and 20 words long have been supplied.

Table 5. Chain length.

	LLC	LOB
range	2-3990 words (0-753 tone units)	2-73 words
2-20 words	43%	73%

Table 5 shows that the chains in LLC were longer than those in LOB. The longest chain in LLC was 753 tone units long, equalling about 3990 words. The longest chain in LOB was 73 words. Moreover, 43% of the chains in LLC were between two and 20 words long, whereas 73% of the chains in LOB had this length.

It was thus shown that pronominal chains in conversation have a tendency to be longer and contain more pronoun tokens than chains in writing. To a large extent this can be explained by factors in the context of situation, by verbal and non-verbal clues, by intimate relationships between the speakers, and by the extent of shared knowledge between the speakers.

CONCLUDING REMARKS

The study of pronominal chains largely confirmed the hypotheses I set up in the introduction. It should, however, be emphasized that the written sample was very small and consisted of informative prose only. Care should thus be taken not to generalize too far. It is also important to remember that this study is in many ways a simplification. No comparison has been made between the individual texts, nor between individual speakers in the conversations.

Still, there are important differences to account for in the conversation corpus both between texts and speakers, due to contextual factors (cf Thavenius 1983). Nonetheless, the tendencies that were shown in this study were interesting. The chain frequencies were much higher in the conversations than in the written texts. There were twice as many chains per 10,000 words in LLC as in LOB, which showed a general tendency to use chains more frequently in conversation than in writing.

It was also shown that, although exophoric chains are uncommon in conversation, they are still more uncommon in writing. In my written material there was in fact not one chain with an exophoric referent. This can be explained by the limited possibilities there are in writing to refer outside the text itself. In conversation it is possible to refer both deictically to the physical situation, and to the abstract context of situation.

The structure of the chains, that is the number of pronouns per chain and the chain length was also studied. The results showed for both corpora that chains containing few tokens were favoured, but that they contained fewer tokens in writing than in conversation. It was also found that chain length was closely related to the number of pronouns per chain, and that the range for chain length was much larger in LLC than in LOB. The proportion of chains that were not longer than 20 words was much higher in LOB (73%) than in LLC (43%). That chains were, on the whole, short, and contained few pronouns, can presumably be related to the fact that each referent is referred to only for a short time. The topic is thus of decisive importance. That chains are longer in conversation than in writing can be explained by factors such as relationship between speakers and the range of their shared knowledge. The context of a conversation also helps to strengthen the chains in supplying clues to the understanding of their meaning.

To sum up: The referential strategy differs between conversation and writing and causes differences in the use of pronominal chains in the two channels. A speaker in a conversation makes pragmatic choices and has no real wish to vary his words. A writer who makes stylistic choices aims at variation in his production, and this in fact acts as a constraint on the use of pronominal chains.

NOTES

I am grateful to Anna-Brita Stenström, University of Lund, and Gunnel Tottie, University of Uppsala, for helpful comments on an earlier draft of this paper.

1. • end of tone unit
   + + simultaneous speech
   For complete prosodic notation see Svartvik & Quirk 1980.

2. For a detailed description of the corpus see Svartvik & Quirk 1980.

3. S.1.1, S.1.3, S.1.4, S.1.5, S.1.6, S.1.8, S.3.3, S.3.4.

4. B03, C13, F27, G36, J26.

REFERENCES

Crystal, D. 1976. Prosodic systems and intonation in English. Cambridge: Cambridge University Press.

Halliday, M.A.K. 1980. "Context of situation". Sophia linguistica. Working papers in linguistics 6.

Kroch, A.S. & D. Hindle. 1980. Progress report to the N.I.E. on grant no. G78-0169,"A quantitative study of the syntax of speech and writing" for the period Jan. 1., 1979 to Jan. 1., 1980. Philadelphia. Mimeo.

Quirk, R., A.P. Duckworth, J. Svartvik, J.P.L. Rusiecki & A.J.T. Colin. 1964. "Studies in the correspondence of prosodic to grammatical features in English". Proceedings of the Ninth International Congress of Linguists. 679-691. The Hague: Mouton.

Svartvik, J. & R. Quirk (eds.) 1980. A corpus of English conversation. Lund studies in English 56. Lund: Gleerup/Liber.

Thavenius, C. 1982. "Exophora in English conversation". Impromptu speech: A symposium, ed. by N.E. Enkvist. Åbo: Åbo Akademi.

Thavenius, C. 1983. Referential pronouns in English conversation. Lund studies in English 64. Lund: Gleerup/Liber.